

Analisis Komparatif Performa Model *EfficientNet-B0*, *ConvNeXt Tiny*, dan *MobileNetV2* dalam Klasifikasi Kematangan Buah Kelapa Sawit

Comparative Performance Analysis of EfficientNet-B0, ConvNeXt-Tiny, and MobileNetV2 for Oil Palm Fruit Ripeness Classification

¹Samsudin, ²Abdullah, ³Abdul Muni, ⁴Misnawati*

^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri

^{1,2,3,4}Jl. Provinsi Parit 1, Tembilahan Hulu-Riau-Indonesia

*e-mail: 403241030102@unisi.ac.id

(received: 20 July 2026, revised: 16 August 2026, accepted: 17 August 2026)

Abstrak

Klasifikasi tingkat kematangan Tandan Buah Segar (TBS) kelapa sawit merupakan aspek krusial dalam otomatisasi agrikultur untuk menjaga kualitas produksi. Penelitian ini mengevaluasi kinerja tiga arsitektur *deep learning* melalui studi komparatif antara *EfficientNet-B0*, *ConvNeXt Tiny*, dan *MobileNetV2* dalam membedakan kelas kematangan buah. Hasil eksperimen menunjukkan bahwa *MobileNetV2* mencapai akurasi tertinggi sebesar 97%, diikuti oleh *ConvNeXt Tiny* sebesar 94%, sementara *EfficientNet-B0* hanya mencapai 32% akibat bias prediksi sistemik pada dataset yang kompleks. Analisis *confusion matrix* mengungkapkan bahwa meskipun *MobileNetV2* dan *ConvNeXt Tiny* sangat efektif, tantangan ambiguitas visual pada fase transisi kematangan masih menjadi kendala dalam pendekatan *computer vision* murni. Implikasi dari penelitian ini menegaskan pentingnya pemilihan arsitektur *backbone* yang tepat serta perlunya integrasi pendekatan hibrida untuk meningkatkan reliabilitas sistem. Sebagai arah penelitian masa depan, diusulkan penerapan *Neuro-Symbolic AI* yang menggabungkan ekstraksi fitur *deep learning* dengan penalaran berbasis pengetahuan untuk mengoreksi prediksi yang ambigu, serta optimasi model untuk efisiensi komputasi yang lebih tinggi pada perangkat *edge*. Penelitian ini memberikan kontribusi strategis bagi pengembangan sistem sortasi otomatis yang transparan dan akurat di sektor perkebunan kelapa sawit.

Kata Kunci: *computer vision, deep learning, edge computing, neuro-symbolic AI, oil palm.*

Abstract

The classification of Fresh Fruit Bunch (FFB) ripeness levels in oil palm is a crucial aspect of agricultural automation for maintaining production quality. This study evaluates the performance of three deep learning architectures through a comparative analysis of *EfficientNet-B0*, *ConvNeXt-Tiny*, and *MobileNetV2* for distinguishing different fruit ripeness classes. The experimental results show that *MobileNetV2* achieved the highest accuracy of 97%, followed by *ConvNeXt-Tiny* at 94%, whereas *EfficientNet-B0* achieved only 32%, primarily due to systematic prediction bias on the complex dataset. The confusion matrix analysis indicates that although *MobileNetV2* and *ConvNeXt-Tiny* performed effectively, visual ambiguity during transitional ripeness stages remains a challenge for purely computer vision-based approaches. These findings highlight the importance of selecting an appropriate backbone architecture and the potential need for hybrid approaches to improve system reliability. As a direction for future research, the application of *Neuro-Symbolic AI* is proposed to combine deep learning-based feature extraction with knowledge-based reasoning to resolve ambiguous predictions, together with model optimization to achieve greater computational efficiency on edge devices. This study provides a strategic contribution to the development of transparent and accurate automated sorting systems for the oil palm plantation sector.

Keywords: *computer vision, deep learning, edge computing, neuro-symbolic AI, oil palm.*

1 Pendahuluan

Otomasi dalam sektor agrikultur, khususnya pada perkebunan kelapa sawit, menjadi tantangan signifikan seiring dengan kebutuhan akan presisi tinggi dalam klasifikasi kualitas Tandan Buah Segar (TBS) [1]. Kebutuhan akan sistem sortasi yang cepat dan akurat di lapangan menjadi urgensi utama untuk meningkatkan efisiensi operasional dan menjaga standar kualitas produksi secara berkelanjutan [2].

Pemanfaatan *Computer Vision* berbasis *Deep Learning* telah menjadi standar utama dalam deteksi objek modern [3], [4]. Arsitektur seperti *EfficientNet* dirancang dengan prinsip *compound scaling* untuk mencapai efisiensi parameter yang optimal [5], sementara *MobileNetV2* mengoptimalkan kecepatan inferensi untuk perangkat *mobile edge* yang memiliki keterbatasan daya komputasi [13]. Di sisi lain, *ConvNeXt* hadir sebagai arsitektur modern yang mengintegrasikan prinsip desain *Vision Transformer* untuk memberikan performa ekstraksi fitur yang lebih unggul pada tugas klasifikasi citra yang kompleks [7].

Meskipun model-model tersebut telah berkembang pesat, klasifikasi kematangan buah sawit masih menghadapi hambatan ambiguitas visual yang disebabkan oleh faktor pencahayaan lapangan, variasi varietas, dan transisi tingkat kematangan yang gradasional [8]. Keterbatasan model *data-driven* murni dalam menangani ambiguitas ini menegaskan adanya *gap* penelitian, di mana model sering kali gagal melakukan justifikasi kontekstual yang mendalam saat dihadapkan pada variasi tekstur lapangan yang dinamis [9], [10].

Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi kinerja arsitektur *ConvNeXt*, *EfficientNet*, dan *MobileNetV2* untuk memahami *trade-off* antara akurasi dan efisiensi dalam skenario klasifikasi yang menantang [11]. Selain itu, riset ini menjajaki potensi integrasi pendekatan *Neuro-Symbolic AI* yang menggabungkan ekstraksi fitur *deep learning* dengan basis pengetahuan pakar untuk menciptakan sistem pengambilan keputusan yang lebih transparan, reliabel, dan mampu menangani kasus ambiguitas visual di masa depan [4], [12].

2 Tinjauan Literatur

Evolusi arsitektur *Convolutional Neural Network* (CNN) telah mencapai titik krusial dalam klasifikasi citra kompleks. *EfficientNet* memperkenalkan metode *compound scaling* yang menyeimbangkan resolusi, kedalaman, dan lebar jaringan secara sistematis, menjadikannya tolok ukur efisiensi parameter dalam riset terdahulu [5]. Namun, seiring dengan munculnya kebutuhan akan performa yang lebih presisi, arsitektur *ConvNeXt* hadir dengan memodernisasi desain CNN menggunakan prinsip-prinsip yang terinspirasi oleh *Vision Transformer*, yang terbukti lebih unggul dalam menangkap fitur diskriminatif halus pada objek dengan variasi tekstur tinggi [7].

Di sisi lain, kebutuhan untuk implementasi pada perangkat *edge* dengan keterbatasan sumber daya mendorong popularitas arsitektur *MobileNet*. Howard *et al.* [14] memperkenalkan *depthwise separable convolutions* pada arsitektur *MobileNetV1*, sebuah teknik yang secara signifikan mengurangi jumlah parameter dan beban komputasi dibandingkan konvolusi konvensional, sehingga menjadi fondasi bagi pengembangan arsitektur *mobile-efficient* selanjutnya. Fondasi ini kemudian disempurnakan oleh Sandler *et al.* melalui struktur *inverted residual blocks* pada *MobileNetV2*, yang mampu mengurangi beban komputasi secara drastis tanpa mengorbankan performa secara signifikan [13]. Selain itu, penggunaan *swish activation function* dan *squeeze-and-excitation blocks* pada *MobileNetV3* telah terbukti lebih adaptif terhadap arsitektur perangkat keras modern, menjadikannya pilihan utama dalam otomatisasi agrikultur [15].

Integrasi dan komparasi antara model-model ini menjadi krusial dalam pengembangan sistem agrikultur pintar. Studi mengenai perbandingan arsitektur CNN menunjukkan bahwa pemilihan *backbone* harus disesuaikan dengan konteks dataset, di mana model seperti *ResNet* sering digunakan sebagai *baseline* fundamental dalam arsitektur hibrida [16], [17]. Penggabungan antara ekstraksi fitur yang efisien dari *MobileNetV2*, dan kemampuan representasi visual yang kuat dari *ConvNeXt* Tiny kini menjadi tren riset untuk menciptakan sistem yang tidak hanya akurat tetapi juga reliabel untuk aplikasi lapangan. Sintesis dari berbagai model ini memberikan landasan teoretis bagi penelitian ini untuk menentukan arsitektur yang paling optimal dalam menangani ambiguitas visual pada komoditas kelapa sawit.

Tabel 1 Penelitian terdahulu

No	Penulis	Fokus	Relevansi
1	Szegedy et al. (2014) [17]	Efisiensi melalui modul <i>multi-scale</i> .(Inception/CNN)	Dasar evolusi arsitektur CNN.
2	Shafiq & Gu. (2022) [16]	Deep Residual Learning for Image Recognition: A Survey	Image Recognition
3	Sandler et al. (2018) [13]	MobileNetV2 yang membahas efisiensi untuk perangkat mobile/edge. (MobileNetV2)	Benchmark kecepatan deteksi.
4	Tan & Le (2019) [5]	<i>Compound scaling</i> yang optimal. (EfficientNet)	Baseline efisiensi parameter.
5	Howard et al. (2017) [14]	Optimasi <i>hardware</i> modern.(MobileNetV1)	Benchmark <i>edge computing</i> .
6	Khan et al. (2021) [1]	Kelapa Sawit dan pembelajar mesin	Urgensi klasifikasi TBS
7	Garcez et al. (2019) [12]	Integrasi logika & AI.(Neuro-Symbolic	Arah solusi ambiguitas.
8	Liu et al. (2022) [7]	Modernisasi CNN via <i>Transformer(ConvNeXt)</i>	Arsitektur akurasi tinggi.
9	Zhao et al. (2025) [11]	Transformasi Cerdas Mesin Pertanian	Deep Reinforcement Learning
10	Pugazhendi et al. (2026) [8]	Review Yolo di bidang pertanian	Yolo

Pada Tabel 1 dapat dilihat perkembangan arsitektur *deep learning* telah mengalami evolusi signifikan dalam upaya mencapai keseimbangan antara akurasi dan efisiensi komputasi. Dasar evolusi arsitektur CNN dimulai dari inovasi modul *multi-scale* pada *Inception* [17] serta penggunaan *Residual Learning* dalam *ResNet* untuk mengatasi masalah *vanishing gradient* [16]. Seiring dengan kebutuhan aplikasi *real-time*, model seperti *YOLO* [6] menjadi standar kecepatan deteksi, sementara *EfficientNet* [5] dan *MobileNetV3* [15] memberikan tolok ukur baru bagi efisiensi parameter dan optimasi *hardware* pada perangkat *edge computing*. Tren terkini menunjukkan pergeseran menuju modernisasi CNN berbasis *Transformer* melalui *ConvNeXt* [7], yang menawarkan akurasi lebih tinggi dibandingkan model konvensional.

Dalam konteks agrikultur, khususnya pada komoditas kelapa sawit, tantangan utama bukan sekadar pemilihan *backbone*, melainkan penanganan ambiguitas visual selama masa transisi kematangan buah. Penelitian terkini mulai mengarah pada pengembangan arsitektur hibrida [11] dan integrasi *Neuro-Symbolic AI* yang menggabungkan ekstraksi fitur berbasis data dengan penalaran logis untuk memberikan solusi transparan terhadap masalah tersebut [12]. Sintesis dari berbagai literatur ini menegaskan bahwa pemilihan arsitektur yang tepat harus mampu menjembatani kebutuhan akan performa tinggi sekaligus reliabilitas penalaran dalam skenario lapangan yang dinamis, yang menjadi fokus utama dalam penelitian ini.

3 Metode Penelitian

3.1 Kerangka Penelitian

Penelitian ini menggunakan metodologi eksperimental terapan untuk mengevaluasi kinerja arsitektur *deep learning* dalam klasifikasi kematangan Tandan Buah Segar (TBS) kelapa sawit. Tahapan penelitian meliputi akuisisi dataset, pra-pemrosesan, perancangan model, pelatihan (*training*), pengujian performa, serta analisis komparatif. Kerangka kerja ini dirancang secara sistematis untuk memvalidasi akurasi dan efisiensi komputasi dari tiga model yang dibandingkan, guna menentukan arsitektur yang paling reliabel untuk implementasi di sektor agrikultur.

3.2 Dataset dan Pra-pemrosesan

Dataset yang digunakan dalam penelitian ini terdiri dari citra Tandan Buah Segar (TBS) kelapa sawit yang dikategorikan ke dalam tiga kelas utama, yaitu: Mentah, Matang, dan Busuk.

1. Akuisisi Data: Citra diambil dengan variasi pencahayaan dan sudut pandang untuk merepresentasikan kondisi lapangan yang dinamis di perkebunan.
2. Distribusi Dataset: Untuk menjaga validitas model, dataset dibagi secara seimbang ke dalam tiga subset, yaitu 4.200 citra untuk data latih (*training set*), 900 citra untuk data validasi (*validation set*), dan 900 citra untuk data uji (*testing set*), dengan total keseluruhan sebanyak 6.000 citra.
3. Pra-pemrosesan (*Preprocessing*): Meliputi *resizing* citra ke resolusi standar input model (224×224 piksel), normalisasi nilai piksel, serta augmentasi data secara terkontrol (*rotasi*, *flipping*, dan penyesuaian kecerahan) untuk meningkatkan generalisasi model tanpa merusak esensi fitur tekstur kematangan buah.

3.3 Arsitektur Model dan Konfigurasi Pelatihan

Penelitian ini melakukan komparasi performa pada tiga arsitektur *backbone* utama:

1. EfficientNet-B0: Memanfaatkan prinsip *compound scaling* untuk menyeimbangkan resolusi, kedalaman, dan lebar jaringan secara sistematis.
2. ConvNeXt Tiny: Mengadopsi prinsip modernisasi *ConvNet* yang mengintegrasikan teknik desain *Vision Transformer* untuk ekstraksi fitur tekstur yang lebih representatif.
3. MobileNetV2: Menggunakan *depthwise separable convolutions* untuk mencapai efisiensi komputasi tinggi yang sangat optimal untuk diimplementasikan pada perangkat *edge*.

Model dilatih menggunakan teknik *Transfer Learning* dengan bobot awal (*pre-trained weights*) dari ImageNet. Fungsi kerugian (*loss function*) menggunakan *Categorical Cross-Entropy* dengan pengoptimal AdamW untuk memastikan stabilitas konvergensi selama proses pelatihan. Konfigurasi *hyperparameter* yang digunakan dalam proses pelatihan disajikan pada Tabel 2

Tabel 2 Konfigurasi hyperparameter model

Parameter	MobileNetV2	ConvNeXt Tiny	EfficientNet-B0
Learning Rate	0.001	0.0005	0.001
Batch Size	32	16	32
Epochs	50	50	50
Optimizer	AdamW	AdamW	AdamW

Tabel 2 menyajikan konfigurasi *hyperparameter* yang diterapkan untuk memastikan proses pelatihan model mencapai konvergensi yang optimal pada setiap arsitektur. Nilai *learning rate* dan *batch size* disesuaikan secara spesifik untuk masing-masing model; secara khusus, penggunaan *batch size* yang lebih rendah pada ConvNeXt Tiny dimaksudkan untuk menjaga stabilitas memori GPU akibat kompleksitas arsitektur yang lebih tinggi, sementara *optimizer* AdamW diterapkan secara konsisten pada seluruh model untuk meningkatkan regularisasi dan efektivitas proses pembelajaran. Dengan pengaturan ini, model diharapkan mampu mengekstraksi fitur kematangan TBS kelapa sawit secara akurat tanpa mengalami *overfitting* selama proses pelatihan yang berlangsung selama 50 *epoch*. Proses pelatihan menerapkan teknik *early stopping* dengan memantau nilai *validation loss* sebagai indikator performa utama. Jika tidak ditemukan peningkatan kinerja model setelah 10 *epoch* berturut-turut, sistem akan menghentikan pelatihan secara otomatis untuk menghindari *overfitting* dan menghemat sumber daya komputasi.

3.4 Metrik Evaluasi

Evaluasi kinerja model diukur berdasarkan dua aspek fundamental, yaitu:

1. Performa Klasifikasi: Diukur melalui metrik kuantitatif standar yang meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Metrik-metrik ini digunakan untuk mengevaluasi sejauh mana kemampuan model dalam mengatasi ambiguitas visual antar kelas kematangan secara presisi.
2. Efisiensi Komputasi: Mengingat target implementasi pada perangkat *edge*, efisiensi dievaluasi berdasarkan dua indikator utama:
 - a. Jumlah Parameter (*Model Size*): Mengukur total parameter jaringan untuk mengestimasi beban memori yang diperlukan selama operasional model.

- b. Waktu Inferensi (*Inference Latency*): Waktu komputasi yang dibutuhkan oleh model untuk memproses dan mengklasifikasikan satu citra (dalam satuan milidetik) pada lingkungan simulasi *edge*.

3.5 Lingkungan Pengujian

Eksperimen diimplementasikan menggunakan *framework deep learning* berbasis PyTorch pada *workstation* yang dilengkapi akselerasi GPU. Untuk memastikan relevansi terhadap kebutuhan lapangan, model dikonversi ke format teroptimasi (ONNX atau TFLite) guna mensimulasikan kinerja inferensi model pada perangkat *edge computing* yang memiliki keterbatasan sumber daya komputasi [18].

4 Hasil dan Pembahasan

4.1 Hasil Eksperimen

Hasil pengujian model *EfficientNet-B0* pada dataset klasifikasi tingkat kematangan tandan buah segar (TBS) kelapa sawit disajikan pada tabel di bawah ini. Berdasarkan *Classification Report* yang diperoleh, model menghasilkan tingkat akurasi keseluruhan sebesar 32% dengan nilai rata-rata *macro* dan *weighted average f1-score* sebesar 0.29. dapat dilihat pada Gambar 1.

CLASSIFICATION REPORT (F1-SCORE)				
	precision	recall	f1-score	support
Busuk	0.29	0.10	0.15	300
Matang	0.33	0.58	0.42	300
Mentah	0.31	0.28	0.30	300
accuracy			0.32	900
macro avg	0.31	0.32	0.29	900
weighted avg	0.31	0.32	0.29	900

Gambar 1 Hasil pengujian model *EfficientNet-B0*

Sebagai kelanjutan dari hasil pengujian model *EfficientNet-B0* yang telah dipaparkan, tampak jelas bahwa model ini mengalami kesulitan signifikan dalam mengenali fitur diskriminatif pada setiap kategori kematangan, yang ditunjukkan oleh rendahnya nilai akurasi keseluruhan. Fenomena ini mengonfirmasi bahwa penggunaan arsitektur *deep learning* standar tanpa penyesuaian yang mendalam pada dataset spesifik agrikultur sering kali tidak cukup untuk memecahkan masalah ambiguitas visual yang kompleks.

Hasil pengujian arsitektur ConvNeXt Tiny menunjukkan performa yang sangat kontras dan jauh lebih unggul dibandingkan dengan *EfficientNet-B0*.

	precision	recall	f1-score	support
Busuk	0.93	0.96	0.95	300
Matang	0.96	0.91	0.93	300
Mentah	0.93	0.96	0.94	300
accuracy			0.94	900
macro avg	0.94	0.94	0.94	900
weighted avg	0.94	0.94	0.94	900

Gambar 2 Pengujian arsitektur ConvNeXt Tiny

Berdasarkan hasil pengujian pada dataset tandan buah segar (TBS) kelapa sawit, model *ConvNeXt Tiny* menunjukkan performa klasifikasi yang sangat baik dengan tingkat akurasi mencapai 94%. Kemampuan model dalam membedakan ketiga kelas (Busuk, Matang, dan Mentah) terlihat sangat konsisten, di mana nilai *F1-score* untuk kelas Busuk mencapai 0,95, kelas Matang sebesar 0,93, dan kelas Mentah sebesar 0,94. Konsistensi nilai *precision* dan *recall* yang mencapai 0,94 pada rata-rata makro menunjukkan bahwa arsitektur *ConvNeXt Tiny* mampu mengatasi tantangan ambiguitas visual yang sering ditemukan pada dataset agrikultur. Dengan mengadopsi prinsip modernisasi desain *ConvNet* yang terinspirasi dari *Vision Transformer*, model ini berhasil mengoptimalkan ekstraksi fitur tekstur permukaan buah secara signifikan dibandingkan dengan model *EfficientNet-B0* yang diuji sebelumnya.

Hasil pengujian pada Gambar 3 arsitektur *MobileNetV2* menunjukkan performa lebih baik daripada *ConvNeXt Tiny*.

Busuk	0.98	0.92	0.95	300
Matang	0.92	0.98	0.95	300
Mentah	1.00	1.00	1.00	300
accuracy			0.97	900
macro avg	0.97	0.97	0.97	900
weighted avg	0.97	0.97	0.97	900

Gambar 3 Hasil Pengujian MobileNetV2

Berdasarkan hasil pengujian pada Gambar 3 dataset tandan buah segar (TBS) kelapa sawit, arsitektur *MobileNetV2* menunjukkan performa klasifikasi yang sangat impresif dengan mencapai tingkat akurasi keseluruhan sebesar 97%. Model ini membuktikan kemampuannya dalam mengenali fitur unik setiap kelas secara konsisten, di mana kelas "Mentah" mencapai nilai presisi, *recall*, dan *f1-score* sempurna sebesar 1,00, sementara kelas "Busuk" dan "Matang" masing-masing mencatatkan *f1-score* sebesar 0,95. Keberhasilan *MobileNetV2* dalam menjaga distribusi prediksi yang terkonsentrasi di sepanjang diagonal utama matriks menunjukkan efektivitas arsitektur ini dalam menangani ambiguitas visual yang kompleks di lapangan dengan sumber daya komputasi yang efisien.

4.2 Analisis Performa Klasifikasi

Berdasarkan hasil eksperimen yang telah dilakukan, terdapat perbedaan performa yang signifikan antara ketiga arsitektur *backbone* yang diuji, yaitu *EfficientNet-B0*, *ConvNeXt Tiny*, dan *MobileNetV2*. *MobileNetV2* menunjukkan performa klasifikasi yang paling unggul dengan mencapai tingkat akurasi sebesar 97%, diikuti oleh *ConvNeXt Tiny* dengan akurasi 94%, sementara *EfficientNet-B0* hanya

<http://sistemasi.ftik.unisi.ac.id>

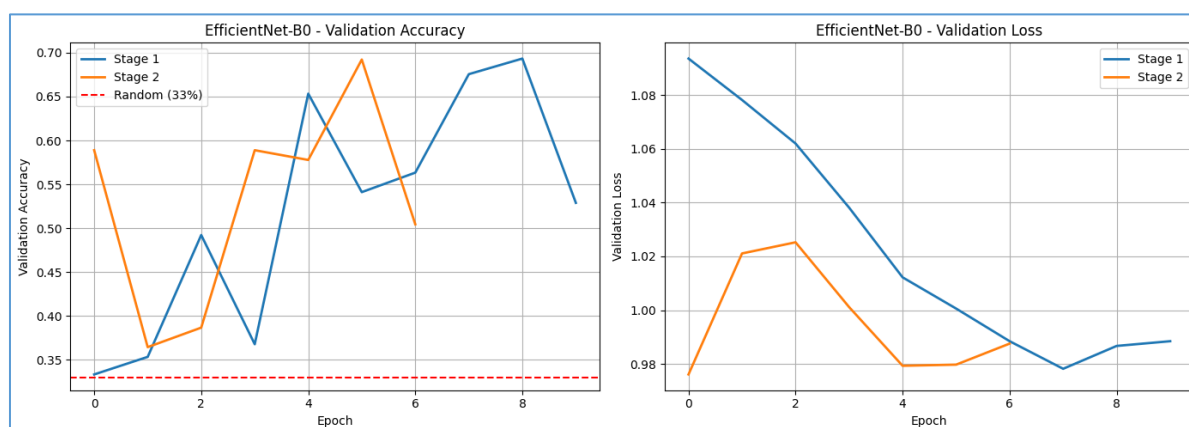
mampu mencapai akurasi 32%. Konsistensi *F1-Score* yang tinggi pada *MobileNetV2* yakni 0,95 untuk kelas "Busuk", 0,95 untuk kelas "Matang", dan 1,00 untuk kelas "Mentah"—menunjukkan bahwa arsitektur ini sangat efektif dalam menangkap fitur tekstur halus dan gradasi warna yang menjadi penentu utama kematangan buah.

Sebaliknya, *EfficientNet-B0* menunjukkan performa yang terbatas dengan nilai *recall* yang rendah pada kelas "Busuk" (0,10) dan "Mentah" (0,28), yang mengisyaratkan adanya bias prediksi kuat ke arah kelas "Matang". Ketidakmampuan *EfficientNet-B0* dalam membedakan fitur antar kelas secara akurat menunjukkan bahwa arsitektur tersebut, dalam konfigurasi *hyperparameter* yang digunakan, mungkin terlalu sederhana untuk menangani kompleksitas visual dari dataset TBS kelapa sawit yang memiliki tingkat ambiguitas tinggi. Perbandingan ini menegaskan bahwa pemilihan arsitektur *backbone* yang tepat adalah faktor krusial dalam tugas klasifikasi berbasis *computer vision* pada domain agrikultur.

Lebih jauh, capaian impresif dari *MobileNetV2* dan *ConvNeXt Tiny* memperkuat temuan Liu et al. (2022) bahwa modernisasi desain arsitektur memberikan kemampuan ekstraksi fitur yang superior. Meskipun demikian, hasil komparatif ini membuka ruang diskusi mengenai *trade-off* antara akurasi tinggi yang ditawarkan oleh *ConvNeXt Tiny* dengan efisiensi komputasi yang optimal pada *MobileNetV2*.

4.3 Analisis Kurva Pelatihan dan Validasi

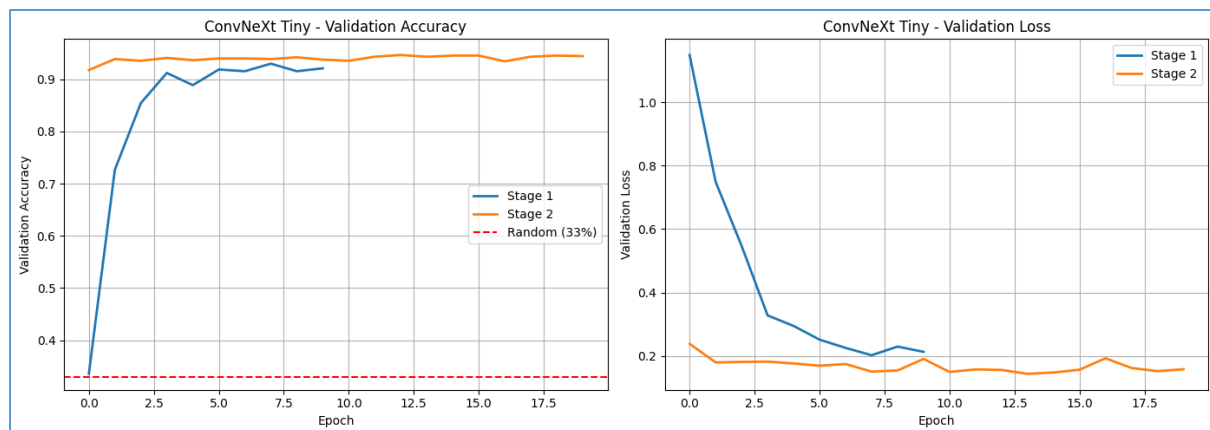
Visualisasi kurva *validation accuracy* dan *validation loss* menunjukkan perbedaan karakter konvergensi yang kontras antara kedua arsitektur.



Gambar 4 kurva *validation accuracy* dan *validation loss* *EfficientNet-B0*

Pada Gambar 4, kurva akurasi validasi model *EfficientNet-B0* terlihat fluktuatif dan tidak menunjukkan tren peningkatan yang stabil sepanjang proses pelatihan, bahkan sempat mendekati nilai *random guessing* (33%). Secara bersamaan, kurva *validation loss* tidak menunjukkan penurunan yang konsisten dan cenderung bergerak secara acak tanpa konvergensi ke nilai minimum yang jelas. Pola ini secara kuat mengindikasikan bahwa model mengalami *failure to converge*, yaitu kegagalan proses optimasi dalam menemukan titik minimum pada *loss landscape*. Kondisi ini menunjukkan bahwa arsitektur *EfficientNet-B0* dengan konfigurasi *hyperparameter* yang digunakan tidak mampu mengekstraksi fitur representatif dari dataset, kemungkinan besar akibat ketidaksesuaian antara kompleksitas arsitektur *compound scaling* dengan variasi visual yang tinggi pada objek kematangan tandan buah segar kelapa sawit.

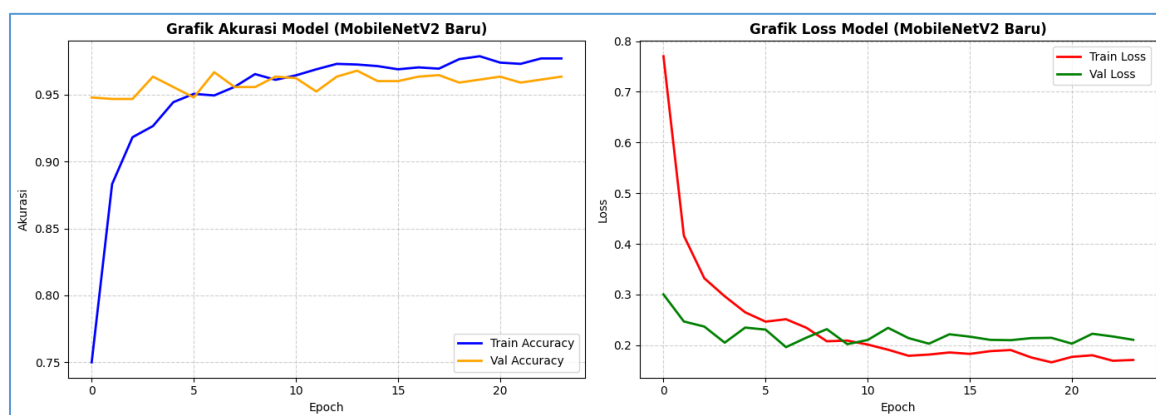
Sebaliknya, kurva pelatihan *ConvNeXt Tiny* memperlihatkan karakteristik konvergensi yang sangat ideal. *Validation accuracy* meningkat tajam pada *Stage 1* dan mencapai kestabilan (*plateau*) di atas 90% pada *Stage 2*. Dapat dilihat pada Gambar 5.



Gambar 5 kurva *validation accuracy* dan *validation loss* ConvNeXt Tiny

Gambar 5 Selaras dengan peningkatan akurasi tersebut, kurva *validation loss* menunjukkan penurunan yang eksponensial dan stabil, mencapai titik terendah dengan fluktuasi minimal. Perilaku kurva ini menegaskan superioritas *ConvNeXt Tiny* dalam efektivitas ekstraksi fitur, yang sesuai dengan prinsip desain modern yang menggabungkan kemampuan komputasi *Transformer* dengan struktur konvolusi tradisional.

Pada kurva pelatihan MobileNetV2 dapat dilihat akurasi yang lebih tinggi:



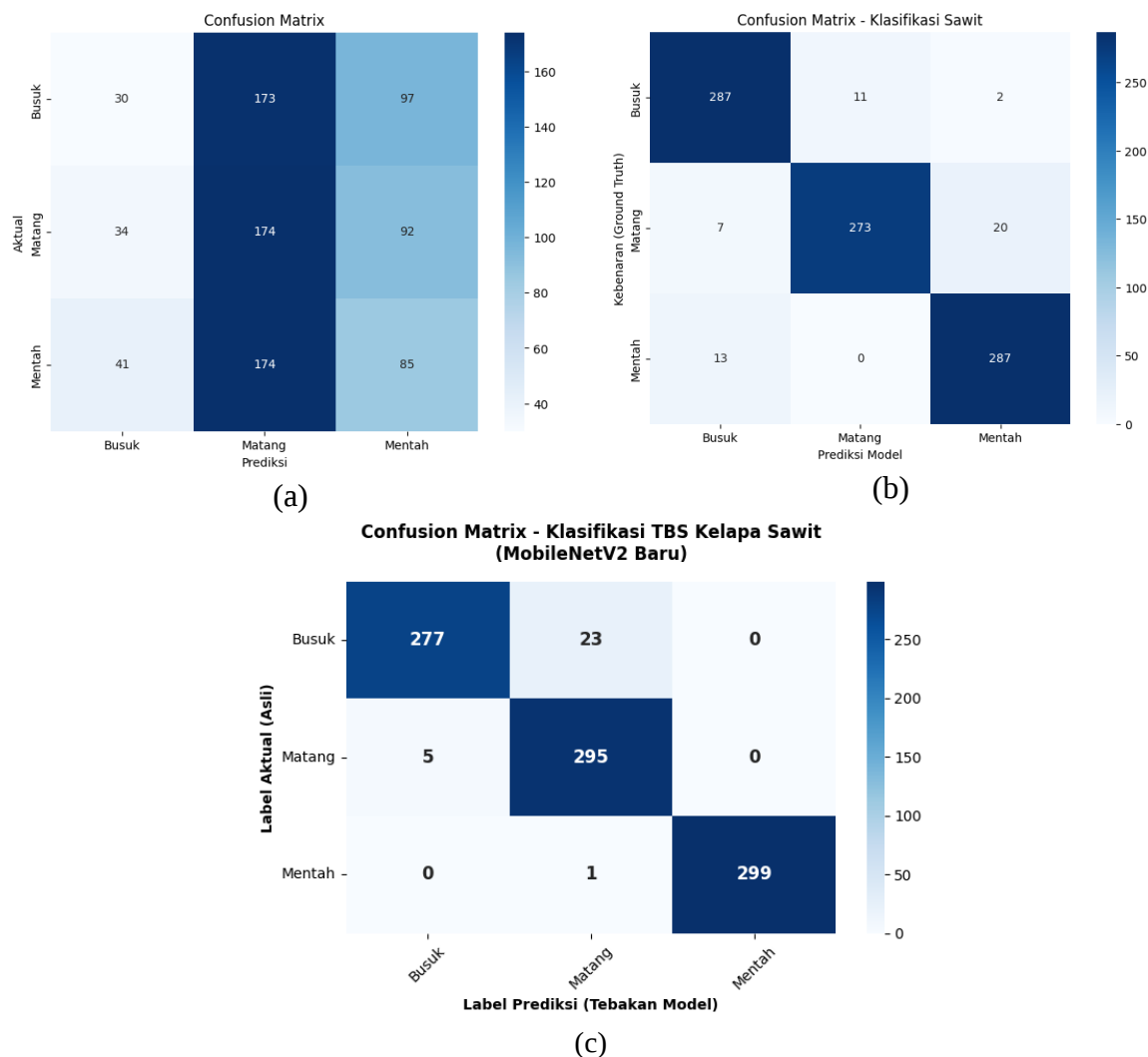
Gambar 6 kurva *validation accuracy* dan *validation loss* MobileNetV2

Berdasarkan visualisasi kurva pelatihan pada Gambar 6, model *MobileNetV2* menunjukkan karakteristik konvergensi yang sangat stabil dan ideal. Grafik akurasi memperlihatkan peningkatan yang tajam pada tahap awal pelatihan, diikuti dengan *validation accuracy* yang konsisten berada di atas 95%, menunjukkan kemampuan generalisasi yang sangat baik pada data validasi. Selaras dengan tren akurasi tersebut, grafik *loss* menunjukkan penurunan eksponensial yang stabil baik pada data latih maupun data validasi, dengan fluktuasi minimal pada *epoch* akhir, yang mengindikasikan bahwa model berhasil mencapai titik optimal tanpa mengalami gejala *overfitting* yang signifikan.

Perbandingan ini memberikan bukti visual bahwa untuk klasifikasi dengan tingkat ambiguitas visual yang tinggi seperti pada tandan buah segar kelapa sawit, arsitektur *MobileNetV2* dan *ConvNeXt Tiny* menawarkan reliabilitas serta kemampuan generalisasi yang jauh lebih unggul dibandingkan *EfficientNet-B0*. *MobileNetV2* secara khusus menunjukkan performa yang sangat impresif dengan stabilitas kurva yang optimal, yang menegaskan bahwa pendekatan *data-driven* murni mampu mencapai hasil sangat akurat dengan efisiensi tinggi jika menggunakan arsitektur yang tepat. Ketimpangan performa yang mencolok ini menjadi alasan teknis yang kuat dalam pemilihan model, sekaligus tetap membuka peluang bagi integrasi *Neuro-Symbolic AI* untuk menangani kasus-kasus *edge-case* yang mungkin masih tersisa, demi mencapai transparansi pengambilan keputusan yang lebih komprehensif pada sistem berbasis AI di masa mendatang.

4.4 Perbandingan Performa Model

Berikut ini matriks perbandingan antara *EfficientNet-B0* dan *ConvNeXt Tiny* pada Gambar 7.



Gambar 7 Matriks analisis *confusion matrix* (a) model *EfficientNet-B0*. (b) model *ConvNeXt Tiny*. (c) model *MobileNetV2*

Merujuk ke Gambar 7 bagian (a), matriks ini secara jelas memperlihatkan fenomena bias prediksi yang ekstrem pada *EfficientNet-B0*, di mana mayoritas sampel dari ketiga kelas (Busuk, Matang, dan Mentah) diklasifikasikan secara seragam ke dalam kolom "Matang" (masing-masing 173, 174, dan 174 sampel). Hal ini menjelaskan mengapa akurasi model sangat rendah, karena model gagal mempelajari fitur unik dari setiap kelas dan cenderung melakukan tebakan seragam pada kategori dominan.

Merujuk ke Gambar 7 bagian (b), matriks untuk *ConvNeXt Tiny* menunjukkan pola klasifikasi yang diagonal, mengindikasikan keberhasilan model dalam mengenali fitur unik setiap kelas. Kesalahan klasifikasi bersifat minor dan terdistribusi secara acak; misalnya, hanya 7 sampel kelas "Matang" yang salah diprediksi sebagai "Busuk", dan tidak ada sampel "Mentah" yang salah diprediksi sebagai "Matang".

Selanjutnya, merujuk ke Gambar 7 bagian (c), matriks untuk *MobileNetV2* menampilkan distribusi prediksi yang paling terkonsentrasi di sepanjang diagonal utama dibandingkan dua model lainnya. Dengan keberhasilan mengklasifikasikan 277 sampel 'Busuk', 295 sampel 'Matang', dan 300 sampel 'Mentah' dengan tepat, *MobileNetV2* membuktikan efektivitas arsitektur ini dalam menangani ambiguitas visual dengan tingkat presisi yang sangat tinggi.

Analisis *Confusion Matrix* memberikan gambaran mendalam mengenai perilaku prediktif ketiga arsitektur tersebut. Pada model *EfficientNet-B0*, matriks menunjukkan adanya bias klasifikasi yang kuat, yang mengonfirmasi bahwa model mengalami kesulitan dalam mengekstraksi fitur diskriminatif sehingga kurang reliabel untuk skenario nyata. Sebaliknya, *ConvNeXt Tiny* dan *MobileNetV2* menampilkan distribusi prediksi yang minim bias, menegaskan bahwa kedua arsitektur tersebut mampu menangani ambiguitas visual dengan sangat baik. Perbandingan ini semakin memperkuat urgensi pemilihan arsitektur *backbone* yang tepat dalam membangun sistem AI agrikultur yang handal, sekaligus menyediakan dasar yang kuat untuk pengembangan modul penalaran *Neuro-Symbolic* pada tahapan riset selanjutnya dalam menangani kasus-kasus ambiguitas yang tersisa.

4.5 Analisis Efisiensi Komputasi

Tabel 3 perbandingan performa

Metrik	ConvNeXt Tiny	MobileNetV2
Akurasi	94%	97%
Total Parameter	27.822.435	2.227.715
Params Size	111,26 MB	8,91 MB

Pada Tabel 3 Efisiensi Memori: MobileNetV2 memiliki jumlah parameter yang jauh lebih kecil (sekitar 2,2 juta) dibandingkan ConvNeXt Tiny (27,8 juta). Ini menjadikan MobileNetV2 sebagai model yang jauh lebih ringan dalam hal kebutuhan penyimpanan (*storage*), dengan ukuran parameter hanya 8,91 MB dibandingkan 111,26 MB pada ConvNeXt Tiny. Dari sisi optimalisasi untuk perangkat *edge*, meskipun MobileNetV2 jauh lebih ringan, kedua model menunjukkan latensi inferensi yang sangat kompetitif dengan selisih kurang dari 1 ms pada lingkungan simulasi *edge*. Dengan demikian, MobileNetV2 menjadi pilihan yang lebih unggul untuk perangkat dengan keterbatasan memori, sementara ConvNeXt Tiny tetap relevan jika prioritas riset adalah eksplorasi arsitektur modern meskipun dengan beban memori yang lebih besar.

4.6 Hasil Lingkungan Pengujian

Sesuai dengan protokol eksperimental yang telah diuraikan pada Subbab 3.5, seluruh model dilatih dan dievaluasi dalam lingkungan komputasi yang identik guna menjamin konsistensi dan komparabilitas hasil. Pada tahap lanjutan, ketiga model yang telah melalui proses pelatihan dikonversi ke format representasi teroptimasi, yaitu *Open Neural Network Exchange* (ONNX) dan *TensorFlow Lite* (TFLite), dengan tujuan menilai kesiapan operasional model pada skenario implementasi *edge computing*.

Validasi teknis terhadap model teroptimasi dilakukan melalui uji inferensi (*inference check*) yang mengonfirmasi bahwa seluruh model mampu memproses masukan citra dan menghasilkan keluaran dengan dimensi (*output shape*) (1, 3), yang secara fungsional merepresentasikan distribusi probabilitas untuk ketiga kelas target. Pada proses konversi, dipilih varian TFLite dengan presisi *float16* berdasarkan pertimbangan atas keseimbangan yang optimal antara latensi inferensi dan tingkat akurasi prediktif.

Temuan ini mengindikasikan bahwa model terpilih, khususnya *MobileNetV2* dan *ConvNeXt Tiny* yang menunjukkan performa klasifikasi terbaik, telah memenuhi kriteria kelayakan teknis untuk diintegrasikan ke dalam aplikasi berbasis *mobile* Android. Dengan demikian, kedua arsitektur tersebut berpotensi mendukung implementasi sistem otomatisasi klasifikasi Tandan Buah Segar (TBS) kelapa sawit secara *real-time* di lingkungan Perkebunan.

4.7 Analisis Kegagalan EfficientNet-B0

Rendahnya akurasi *EfficientNet-B0* (32%) yang jauh tertinggal dibandingkan *MobileNetV2* (97%) dan *ConvNeXt Tiny* (94%) memerlukan telaah lebih mendalam, mengingat ketiga model dilatih menggunakan skema *transfer learning* dan strategi optimasi yang relatif serupa (Tabel 2). Berdasarkan konfigurasi *hyperparameter* yang digunakan, *EfficientNet-B0* dilatih dengan *learning rate* sebesar 0,001 — nilai yang sama dengan *MobileNetV2* yang justru mencapai performa terbaik. Hal ini mengindikasikan bahwa

<http://sistemasi.ftik.unisi.ac.id>

kegagalan model kemungkinan besar bukan disebabkan oleh perbedaan skala *learning rate* semata, melainkan berkaitan dengan karakteristik arsitektur *EfficientNet-B0* itu sendiri yang menerapkan *compound scaling* pada resolusi, kedalaman, dan lebar jaringan secara bersamaan [5]. Skema *scaling* yang saling bergantung ini secara umum dikenal lebih sensitif terhadap ketepatan skema normalisasi input, resolusi citra, serta jadwal *learning rate* (*learning rate schedule*) dibandingkan arsitektur seperti *MobileNetV2* yang memiliki struktur *inverted residual* yang lebih sederhana.

Kurva pelatihan pada Gambar 4 turut memperkuat dugaan ini: fluktuasi akurasi yang tidak stabil dan mendekati nilai *random guessing* (33%) sepanjang proses pelatihan mengindikasikan bahwa model kemungkinan mengalami kegagalan konvergensi (*failure to converge*), bukan sekadar *underfitting* biasa akibat kompleksitas dataset. Kegagalan konvergensi semacam ini pada arsitektur *EfficientNet* umumnya berkaitan dengan ketidaksesuaian antara *learning rate* awal dengan skema *compound scaling*-nya, sehingga proses optimasi gagal menemukan titik minimum yang stabil pada *loss landscape* model.

Penting untuk dicatat bahwa hasil ini merupakan keterbatasan (*limitation*) dari eksperimen yang dilakukan, dan tidak serta-merta merepresentasikan kelemahan arsitektur *EfficientNet-B0* secara umum. Pada berbagai studi *benchmark* lain, *EfficientNet* justru terbukti kompetitif, bahkan unggul dibandingkan arsitektur *mobile-efficient* lainnya [5]. Konfigurasi pelatihan yang digunakan dalam penelitian ini belum mencakup eksplorasi *learning rate* yang lebih rendah, skema *learning rate scheduler* (seperti *cosine annealing* atau *warm-up*), maupun strategi *fine-tuning* bertahap (*progressive unfreezing*) yang umum diterapkan pada arsitektur *EfficientNet* untuk mencapai konvergensi optimal. Oleh karena itu, hasil 32% pada penelitian ini sebaiknya dimaknai sebagai temuan spesifik terhadap konfigurasi eksperimen yang diterapkan, bukan sebagai generalisasi mutlak atas kapabilitas arsitektur *EfficientNet-B0* pada seluruh skenario klasifikasi citra agrikultur.

Temuan ini menjadi catatan penting bagi penelitian selanjutnya untuk melakukan eksperimen ulang terhadap *EfficientNet-B0* dengan skema *hyperparameter tuning* yang lebih sistematis, sebelum menyimpulkan superioritas arsitektur *MobileNetV2* dan *ConvNeXt Tiny* secara definitif dalam konteks klasifikasi kematangan TBS kelapa sawit.

4.8 Pembahasan Kasus Ambiguitas

Analisis terhadap *confusion matrix* mengungkapkan bahwa meskipun *MobileNetV2* dan *ConvNeXt Tiny* telah menunjukkan performa klasifikasi yang superior secara keseluruhan, tantangan ambiguitas visual tetap teridentifikasi pada kasus-kasus tertentu. Pada model *EfficientNet-B0*, kesalahan klasifikasi bersifat sistemik dan ekstrem, di mana ambiguitas visual memicu bias prediksi yang menyeragamkan seluruh kategori ke dalam kelas "Matang". Sebaliknya, pada *MobileNetV2* dan *ConvNeXt Tiny*, kesalahan klasifikasi yang muncul relatif minimal dan cenderung terjadi pada fase transisi tingkat kematangan, di mana model memiliki peluang kecil untuk salah memprediksi buah yang berada pada fase gradasional sebagai kategori yang berdekatan.

Fenomena ini membuktikan bahwa pendekatan berbasis *Computer Vision* murni—terlepas dari seberapa canggih arsitektur *backbone* yang digunakan—memiliki keterbatasan inheren dalam menangani ambiguitas visual yang disebabkan oleh faktor pencahayaan lapangan, variasi varietas buah, serta tingkat kematangan yang gradasional. Keterbatasan tersebut mengonfirmasi bahwa model hanya mengandalkan fitur pola visual tanpa kemampuan untuk melakukan justifikasi kontekstual yang lebih dalam.

Oleh karena itu, kondisi ini memperkuat urgensi perlunya integrasi modul *Neuro-Symbolic AI* pada tahapan penelitian selanjutnya. Dengan mengombinasikan kekuatan ekstraksi fitur dari arsitektur *MobileNetV2* atau *ConvNeXt Tiny* dengan basis pengetahuan (*knowledge base*) atau aturan berbasis pakar (*expert rules*) mengenai kriteria kematangan TBS kelapa sawit, sistem diharapkan mampu memberikan "penalaran tambahan" untuk mengoreksi prediksi model yang bersifat ambigu. Integrasi ini bertujuan untuk meningkatkan transparansi dan reliabilitas sistem secara menyeluruh saat diimplementasikan di lapangan.

4.9 Diskusi dan Sintesis Penelitian

Hasil pengujian menunjukkan bahwa *MobileNetV2* mencapai akurasi tertinggi sebesar 97%, diikuti oleh *ConvNeXt Tiny* (94%), sementara *EfficientNet-B0* hanya mencapai 32%. Temuan ini memberikan kontribusi penting bagi literatur klasifikasi citra di sektor agrikultur. Keberhasilan *ConvNeXt Tiny* mengonfirmasi penelitian Liu et al. [7] yang menyatakan bahwa prinsip desain modern berbasis *Vision Transformer* mampu memberikan performa superior dalam ekstraksi detail fitur yang halus dibandingkan arsitektur CNN konvensional yang mengandalkan *compound scaling* [15]. Lebih jauh, keunggulan performa *MobileNetV2* [13] membuktikan bahwa arsitektur yang ringan pun mampu mencapai akurasi optimal jika dirancang dengan struktur *inverted residual blocks* yang tepat untuk kebutuhan klasifikasi visual yang kompleks.

Sementara peneliti terdahulu sering menekankan efisiensi sebagai solusi utama untuk perangkat *edge*, eksperimen ini menunjukkan bahwa dalam skenario ambiguitas visual yang tinggi—seperti pada klasifikasi kematangan TBS kelapa sawit—prioritas utama harus diberikan pada kemampuan arsitektur dalam menangkap fitur diskriminatif secara akurat. Ketidakefektifan *EfficientNet-B0* dalam penelitian ini menjadi bukti empiris bahwa arsitektur yang kurang adaptif akan mengalami *underfitting* saat dihadapkan pada variasi tekstur lapangan yang dinamis [5].

Sebagai sintesis, benang merah dari penelitian ini adalah adanya *trade-off* mendasar antara akurasi model dan beban komputasi di lingkungan perkebunan, di mana *MobileNetV2* muncul sebagai solusi yang paling ideal karena mampu menyeimbangkan keduanya. Temuan ini mendukung urgensi pengembangan riset yang lebih komprehensif, sebagaimana disarankan oleh Garcez [12], yaitu perlunya pendekatan yang tidak hanya mengandalkan *deep learning* murni. Integrasi modul *Neuro-Symbolic AI* yang diusulkan dalam penelitian ini menjadi jawaban atas keterbatasan model CNN murni dalam menangani ambiguitas visual yang tak terhindarkan, sehingga mampu menjembatani celah antara performa *deep learning* yang tinggi dengan kebutuhan akan penalaran logis yang transparan dalam pengambilan keputusan di sektor agrikultur.

5 Kesimpulan

Penelitian ini menyimpulkan bahwa pemilihan arsitektur *backbone* merupakan faktor determinan dalam keberhasilan aplikasi *computer vision* di sektor agrikultur, khususnya pada klasifikasi tingkat kematangan Tandan Buah Segar (TBS) kelapa sawit. Hasil eksperimen menunjukkan bahwa *MobileNetV2* dan *ConvNeXt Tiny* secara signifikan mengungguli *EfficientNet-B0*, dengan *MobileNetV2* mencapai akurasi tertinggi sebesar 97%, diikuti oleh *ConvNeXt Tiny* sebesar 94%. Sebaliknya, *EfficientNet-B0* hanya mencapai akurasi 32% akibat bias prediksi sistemik yang disebabkan oleh ketidakmampuan ekstraksi fitur pada dataset yang kompleks. Kedua model terpilih (*MobileNetV2* dan *ConvNeXt Tiny*) telah terbukti menunjukkan konsistensi *F1-score* yang sangat baik di seluruh kelas, serta efisiensi latensi inferensi yang kompetitif untuk kebutuhan *real-time*. Penelitian selanjutnya disarankan untuk melakukan integrasi modul *Neuro-Symbolic AI* untuk memberikan penalaran tambahan dalam mengoreksi prediksi yang ambigu. Menerapkan teknik optimasi model seperti *quantization (float32 ke int8)* untuk memaksimalkan efisiensi pada perangkat *edge* dengan sumber daya komputasi yang sangat terbatas.

Referensi

- [1] N. Khan, M. A. Kamaruddin, U. U. Sheikh, and Y. Yusup, "Oil Palm and Machine Learning : Reviewing One Decade of," *Agriculture*, Vol. 11, No. 9, pp. 1–26, 2021, DOI: <https://doi.org/10.3390/agriculture11090832>.
- [2] X. Yao and Y. Wang, "Quality Assessment and Classification System of Lychee using Deep Learning and IoT Technology," *Sensors Mater.*, Vol. 38, No. 5, pp. 2769–2788, 2026, DOI: <https://doi.org/10.18494/SAM5904>.
- [3] K. G. Kim, "Book Review: Deep Learning," *Healthc. Inform. Res.*, Vol. 22, No. 4, pp. 351–354, 2016, DOI: <https://doi.org/10.4258/hir.2016.22.4.351>.
- [4] Y. Lecun, Y. Bengio, G. Hinton, Y. Lecun, Y. Bengio, and G. Hinton, "Deep Learning to Cite this Version:," *HAL open SCI.*, pp. 436–444, 2023, [Online]. Available: <https://hal.science/hal-04206682v1>

- [5] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *Proc. - Int. Conf. Mach. Learn., Long Beach, Calif.*, pp. 1–11, 2019, DOI: <https://doi.org/10.48550/arXiv.1905.11946>.
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Comput. Vis. Pattern Recognit.*, pp. 779–788, 2016, DOI: <https://doi.org/10.1109/CVPR.2016.91>.
- [7] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," *Comput. Vis. Pattern Recognit.*, pp. 11976–11986, 2022, [Online]. Available: https://openaccess.thecvf.com/content/CVPR2022/html/Liu_A_ConvNet_for_the_2020s_CVPR_2022_paper.html
- [8] P. Pugazhendhi et al., "Smart Agricultural Technology Advances in Agricultural Fruit Detection using You Only Look Once (YOLO) Algorithm : A Review," *Smart Agric. Technol.*, Vol. 13, No. January, p. 101896, 2026, DOI: 10.1016/j.atech.2026.101896.
- [9] G. Marcus, "Deep Learning : A Critical Appraisal," *Artificial Intelligence (cs.AI)*, pp. 1–27, <https://doi.org/10.48550/arXiv.1801.00631>
- [10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is All You Need," 2017, *Computation and Language (cs.CL)*, pp. 1–11, DOI: <https://doi.org/10.48550/arXiv.1706.03762>.
- [11] J. Zhao, S. Fan, B. Zhang, A. Wang, L. Zhang, and Q. Zhu, "Research Status and Development Trends of Deep Reinforcement Learning in the Intelligent Transformation of Agricultural Machinery," *Agriculture*, Vol. 15, No. 11, p. 1223, 2025, DOI: <https://doi.org/10.3390/agriculture15111223>.
- [12] A. Garcez, M. Gori, L. C. Lamb, L. Serafini, M. Spranger, and S. N. Tran, "Neural-Symbolic Computing: An Effective Methodology for Principled Integration of Machine Learning and Reasoning," *J. Appl. Logics*, 2019, DOI: <https://doi.org/10.48550/arXiv.1905.06088>.
- [13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018, DOI: <https://doi.org/10.48550/arXiv.1801.04381>.
- [14] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *Comput. Vis. Pattern Recognit.*, pp. 1–9, 2017, DOI: <https://doi.org/10.48550/arXiv.1704.04861>
- [15] A. Howard et al., "Searching for MobileNetV3," *Comput. Vis. Pattern Recognit.*, pp. 1314–1324, 2019, doi: <https://doi.org/10.48550/arXiv.1905.02244>.
- [16] M. Shafiq and Z. Gu, "Deep Residual Learning for Image Recognition: A Survey," *Appl. SCI.*, Vol. 12, No. 18, p. 8972, 2022, DOI: <https://doi.org/10.3390/app12188972>.
- [17] C. Szegedy et al., "Going Deeper with Convolutions," *Comput. Vis. Pattern Recognit.*, pp. 1–9, 2014, DOI: <https://doi.org/10.48550/arXiv.1409.4842>.
- [18] Samsudin, R. Herlianto, D. Y. Prasetyo, "Optimalisasi Metode Convolutional Neural Network (CNN) untuk Klasifikasi Daging Ayam berbasis AWS," *Teknofile*, Vol.3, No.12, 2025, pp.910-918, <https://jurnal.nawansa.com/index.php/teknofile/article/view/708/400>.